Introducción a distribuciones de Probabilidad.

Dr. Servio Interiano

Guatemala febrero 2022

Extractado de:



Grado de Medicina – Universidad de Barcelona
Bioestadística básica, Epidemiologia y Introducción a la Investigación (2016/17)
Begoña Campos – Departamento de Fundamentos Clínicos

"all the world believes it (Gaussian distribution) firmly, because the mathematicians imagine that it is a fact of observation, and the observers that it is a theorem of mathematics"

(Lippmann a Poincaré, 1912)

"todo el mundo cree en ella (la distribución de Gauss) firmemente, porque los matemáticos imaginan que está fundamentada en observaciones de hechos, y los observadores [imaginan] que es un teorema de los matemáticos"

(Lippmann a Poincaré, 1912)

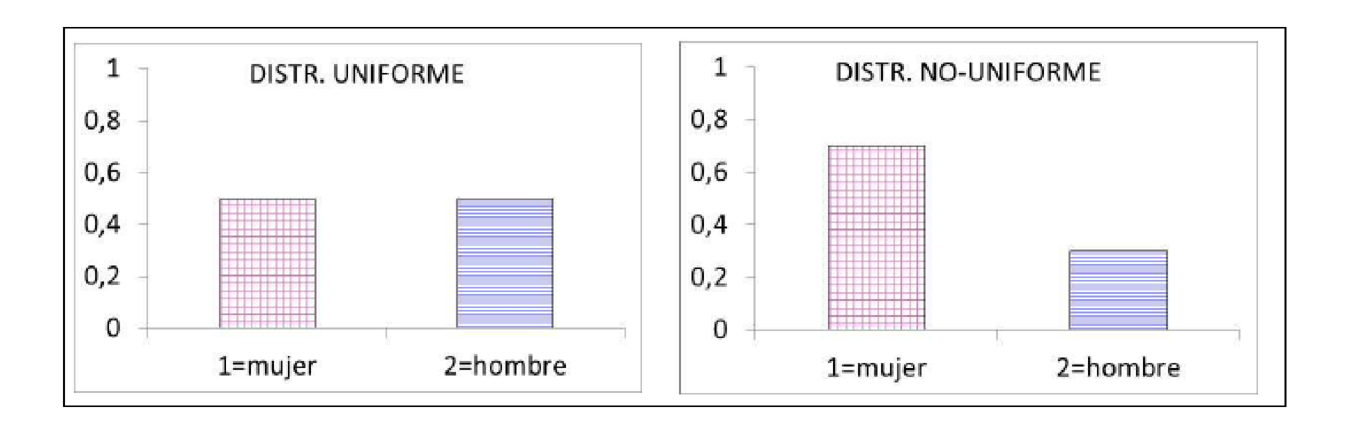
Encyclopedia of Statistical Sciences. Vol 4: Laws of error.

A. El concepto de DISTRIBUCIÓN se refiere al reparto de los individuos de una población según una característica. Supongamos que la característica es el sexo. Si en la población hay tantas mujeres como hombres se dice que la distribución de sexo en la población es uniforme, y por tanto la probabilidad de que un individuo seleccionado al azar sea mujer vale 0,5:

Uniforme: P (M) = 0,5 ; P (H) = 0,5 Suma = 1

Si el reparto no es uniforme, entonces la probabilidad de que el individuo sea mujer valdrá diferente de 0,5, por ejemplo:

No-uniforme: P (M) = 0,7 ; P (H) = 0,3 Suma = 1



B. Sea como sea el reparto, al final la suma de todas las probabilidades ha de ser siempre igual a uno.

C. Hay que distinguir entre distribución de frecuencias y distribución de probabilidad. La primera es un reparto empírico observado en una colección de datos. El histograma es una técnica gráfica para representar la distribución de frecuencias de una variable continua.

Una distribución de probabilidad es un reparto teórico de la población y por tanto es una función matemática. La representación gráfica será una curva como por ejemplo la campana de Gauss.

D. Definición formal. Una VARIABLE ALEATORIA, (**VA**), es una función matemática que asigna diferentes números reales a cada resultado de una experiencia aleatoria.

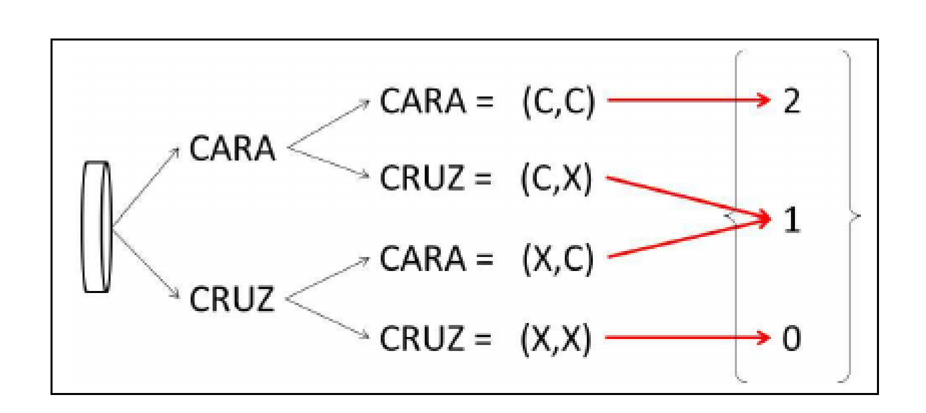
VA: espacio muestral (**S**) → conj. Reales (**R**)

Ejemplo:

E = “lanzar al aire una moneda 2 veces”

S = {(c,c); (c,x); (x,c); (x,x)}

VA = “contar el número de caras”



E. Recorrido de una VA: Conjunto R de todos los valores posibles. En el ejemplo anterior R = {0, 1, 2}.

F. La intención al definir una **Variable Aleatoria (VA)** es poder convertir cualquier espacio muestral en un conjunto de números. Esto se puede hacer de distintas maneras. Ejemplos:

- Asignación arbitraria
  - E = "lanzar una moneda"
    - S = {cara, cruz}
      - VA="valor de la cara obtenida: c=0; x=1"
        - R = {0,1}
- Asignación identidad
  - E = "lanzar un dado de 6 caras"
    - S = {1, 2, 3, 4, 5, 6}
      - VA= "puntos de la cara"
        - R = {1, 2, 3, 4, 5, 6}
- Asignación por medida
  - E = "extraer al azar una persona de una población"
    - S = {Pepe, Luís, Marta, Carlos,......, Ana}
      - VA="altura de una persona"
        - R = {148,152.5, 167.3,...,190.4}

G. Tipos de variables aleatorias

- DISCRETA ("discontinua"). Variable con un conjunto de resultados limitados (recorrido finito o infinito numerable).
  X = número de hijos de una mujer
  Y = número de partos en taxis de Barcelona en un año
- CONTINUA. Variable con un conjunto de resultados ilimitado, es decir, el recorrido es infinito. Ejemplos:
  X = la altura de un individuo
  Y = el salario de un trabajador

H. Una distribución teórica de probabilidad describe el reparto de los valores de una variable aleatoria en una población. Los valores que son más abundantes tendrán mayor probabilidad de aparecer al realizar la experiencia aleatoria que los valores más escasos.

I. El reparto de probabilidad de una variable discreta se define mediante la FUNCIÓN DE PROBABILIDAD que asigna a cada valor posible de la variable aleatoria un número p en el intervalo [0,1]:

FP: $x \rightarrow P(x) = p$ si x está en el recorrido
$x \rightarrow P(x) = 0$ si el valor x no está recorrido

Dos propiedades importantes:

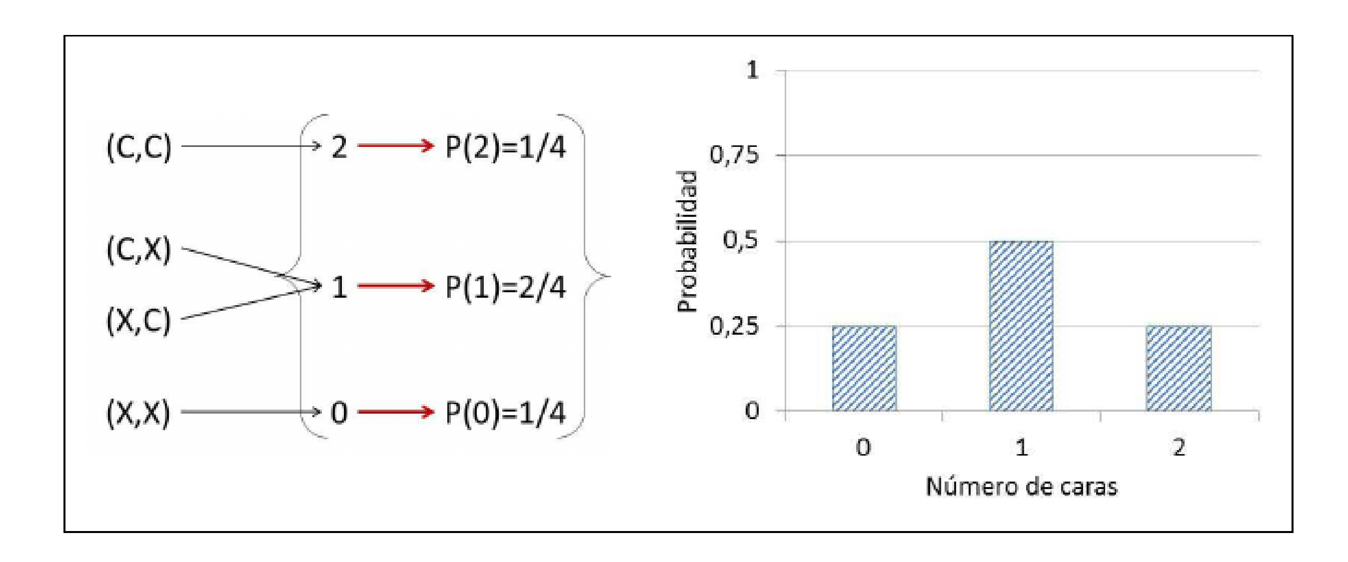
- Los valores p son siempre positivos o nulos: $p \geq 0$
- La suma de todas las p's ha de ser igual a uno: $\Sigma p = 1$

Ejemplo:

E = "lanzar al aire una moneda 2 veces"

S = {(c,c); (c,x); (x,c); (x,x)}

VA = X= "contar el número de caras"



J. En las variables aleatorias continuas no se puede usar la función de probabilidad porque la probabilidad en un punto x, sea cual sea, es siempre cero:

$P(X=x) = 0$

En consecuencia en variables continuas sólo tiene sentido hablar de probabilidad referida a intervalos, que pueden ser muy pequeños:

$P(X \in [a, b]) \neq 0$; siendo a y b números reales cualesquiera

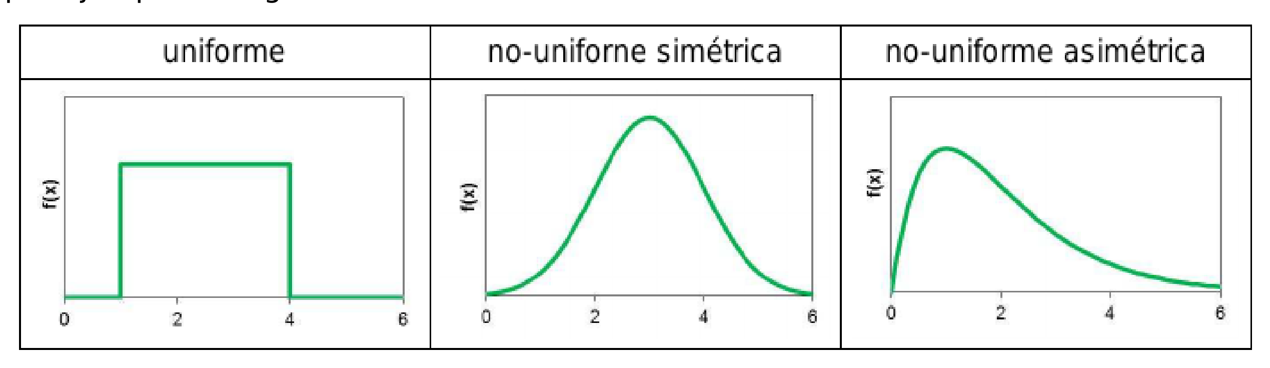
K. Para explicar el reparto de probabilidad de una variable continua se utiliza la FUNCIÓN DE DENSIDAD DE PROBABILIDAD (f) que es una curva continua con las siguientes propiedades:

- En cualquier punto de la recta real la función es positiva o cero:

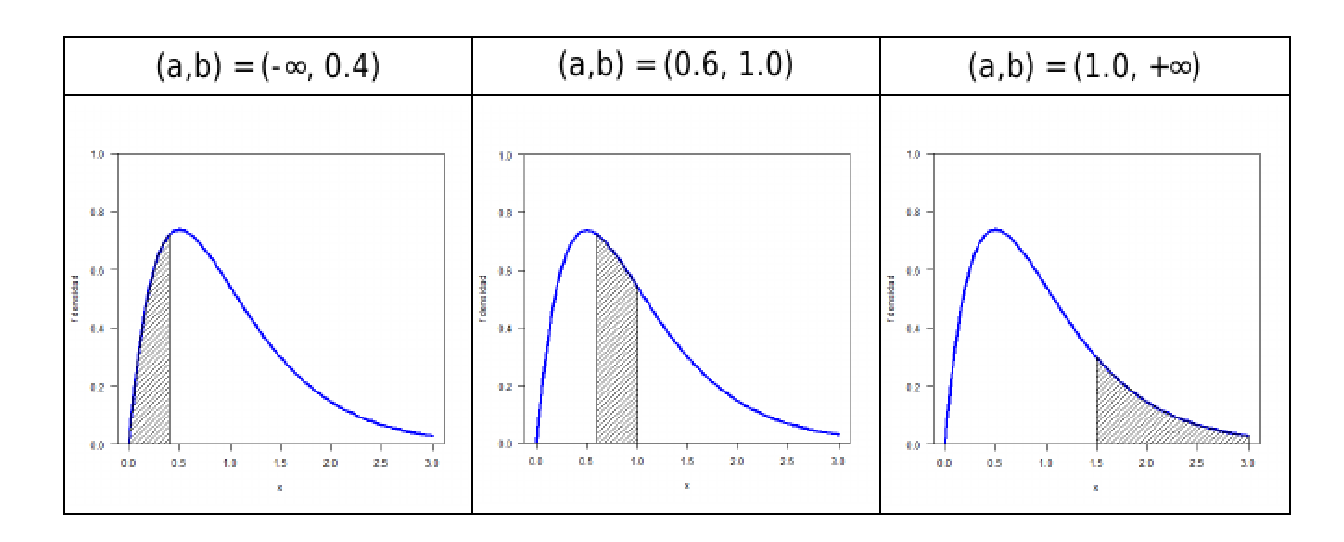
  ≥ 0
- El área total bajo la curva de f vale uno:

  Área total =1

L. Existen muchas curvas que cumplen los criterios para ser función de densidad, por ejemplo las siguientes:



M. La función de densidad de probabilidad servirá para calcular la probabilidad en un intervalo *calculando el área contenida en dicho intervalo*.



Principales Distribuciones de probabilidad.

A. El modelo **BINOMIAL** explica el reparto de probabilidad de los resultados que se pueden observar al realizar una experiencia aleatoria del tipo:

E = “repetir n veces un ensayo con resultado dicotómico”

y la variable aleatoria es el recuento de los éxitos

X = ”número de éxitos en n repeticiones del ensayo”

R = {0,1…, n}

Ejemplo

E = “escoger al azar 3 persona de un grupo” 6 → éxito = “ser zurdo”
S = {(d, d, d); (d, d, z); (d, z, d); (z, d, d); (d, z, z); (z, d, z); (z, z, d); (z, z, z)}
X = ”número de zurdos en 3 personas escogidas al azar”
R = {0, 1, 2, 3}

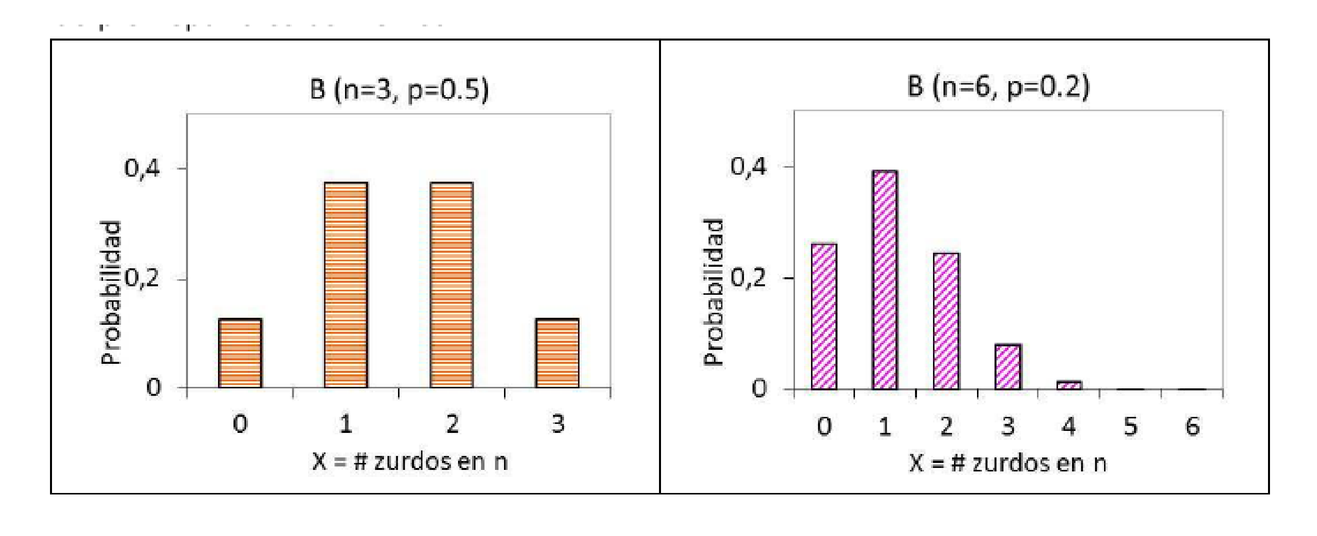
Para decir que una variable aleatoria X sigue una distribución de probabilidad Binomial se utiliza la expresión:

$X \sim B(n,p)$

siendo n y p los dos parámetros del modelo:

n = número de repeticiones
p = probabilidad de éxito (1-p= q =probabilidad de fracaso)

La forma de la función de probabilidad dependerá de los valores de los parámetros. El número de barras es igual a (n+1) y por tanto a mayor n más bajas son las barras.
La simetría del conjunto sólo ocurre para p=0.5, mientras que para cualquier otro valor de p el reparto es asimétrico.



B. El modelo de **POISSON** explica el reparto de probabilidad de una variable aleatoria
que se ajuste a una expresión del tipo:

X = ”número de acontecimientos raros en un intervalo de tiempo o espacio”
R = {0,1…, ∞}

En esencia, se trata de un recuento sin límite superior de resultados con una probabilidad muy baja.

Ejemplos

- número mujeres que dan a luz durante un vuelo de avión en un año
- número de nuevos casos de cáncer de mama en varones en una región

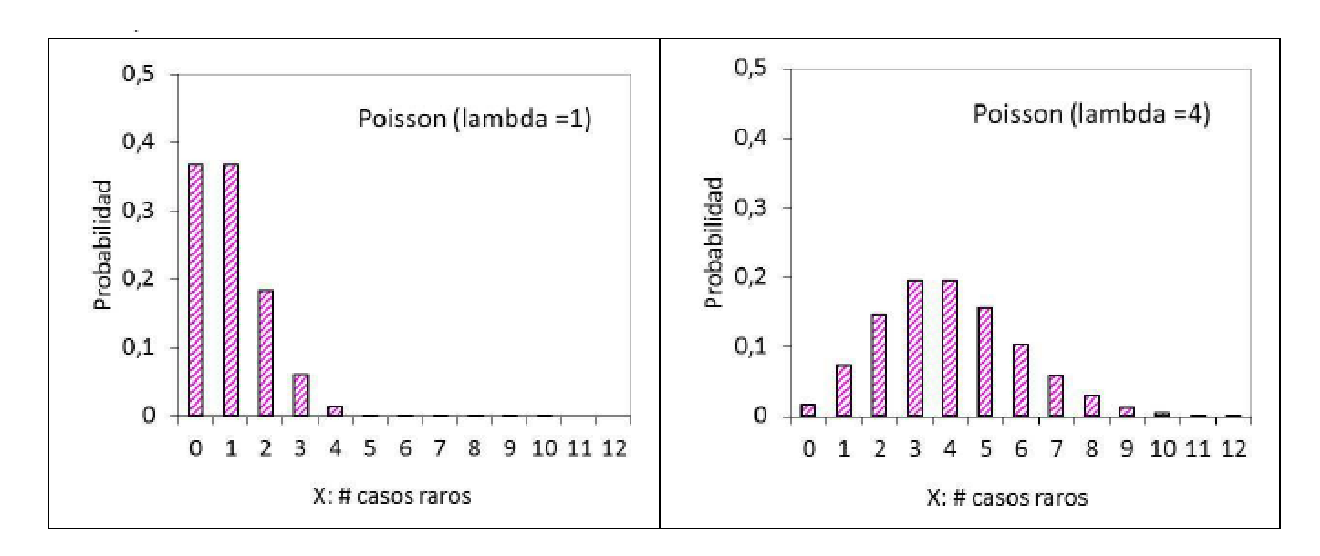
Para decir que una variable aleatoria X sigue una distribución de probabilidad Poisson se utiliza la expresión:

X ~ P (lambda)

siendo lambda (λ) el único parámetro de la distribución

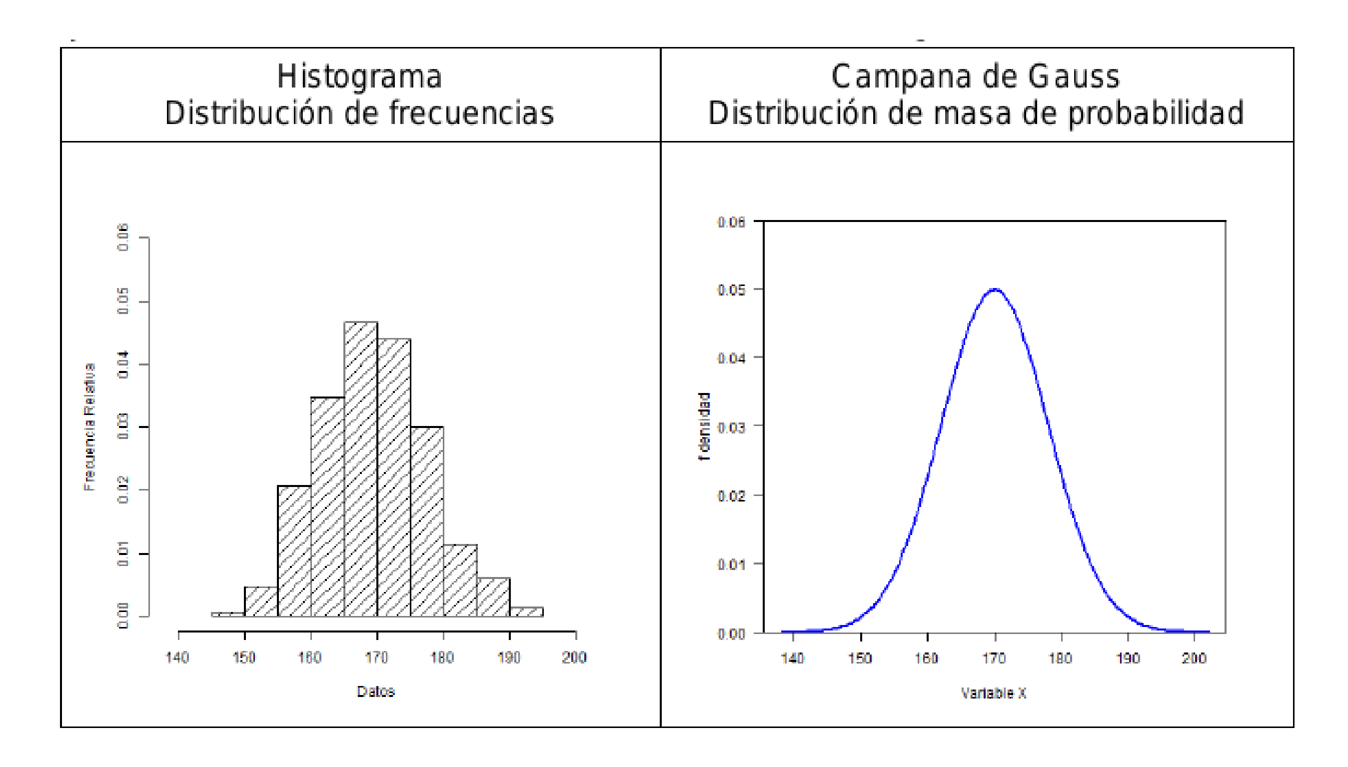
La forma de la función de probabilidad dependerá sólo del valor lambda.

Teóricamente el recorrido de la variable acaba en infinito por eso el número de barras es siempre ilimitado.



C. La distribución **NORMAL** es un objeto matemático. Fue la solución encontrada por A. de Moivre para aproximar cálculos de probabilidad en juegos de azar a principios del siglo XVIII. Posteriormente fue empleada para explicar los errores de medida en física y astronomía. Galileo ya había razonado que los errores de observación se distribuían simétricamente y tendían a agregarse en torno a su valor real.

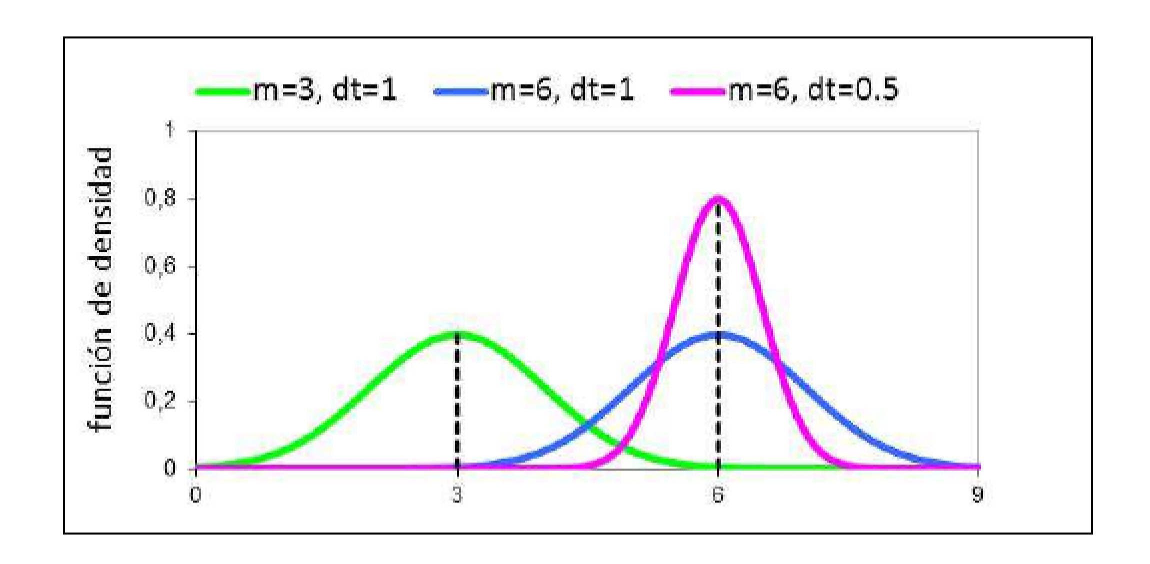
K.F. Gauss la utilizó en un trabajo que publicó en 1809 y su nombre acabó identificando la distribución porque la usó a menudo cuando analizaba datos astronómicos. El protagonismo de esta distribución en las ciencias de la vida se debe al desarrollo de la Biometría. Hacia finales del siglo XIX se generalizó el uso de esta distribución de probabilidad para justificar la distribución de frecuencias de características biológicas.



La campana de Gauss es la representación gráfica de la función de densidad, f, del modelo Normal.

Los parámetros del modelo son μ y σ. El primero, μ, es la esperanza de X y decide la posición de la campana a los largo del eje horizontal. El segundo, σ, es la raíz cuadrada de la variancia X y determina la anchura de la campana.

Parámetros de la distribución normal: (μ, σ)



Características de la densidad normal

- por definición de densidad, el área total bajo la curva vale 1.
- la curva presenta un pico (máximo) en el centro, cuya altura puede ser mayor que 1 si la desviación típica es muy pequeña.
- la curva es simétrica respecto al eje vertical que pasa por el centro.
- la media, mediana y moda coinciden en el mismo punto.
- la función tiene dos puntos de inflexión en las abscisas $(\mu-1\sigma)$ y $(\mu+1\sigma)$ respectivamente, con lo cual la desviación típica es la distancia desde el eje central a los puntos de inflexión.
- las colas son infinitas, es decir, la curva toca el eje X en $-\infty$ y $+\infty$.
- cualquier campana de Gauss cumple siempre las siguientes proporciones:
  - área entre $(\mu-1\sigma)$ y $(\mu+1\sigma) \approx 68\%$
  - área entre $(\mu-2\sigma)$ y $(\mu+2\sigma) \approx 95\%$
  - área entre $(\mu-3\sigma)$ y $(\mu+3\sigma) \approx 99\%$